

УДК 004.738.78:0.25.4.036

ПОКРЕТ ЗА ОТВОРЕНЕ АРХИВЕ (ОАИ)¹

Богољуб Мазих
Универзитетска библиотека "Светозар Марковић"
Београд

Сажетак

Аутор објашњава принципе и дефинише појмове и значај активности на изградњи отворених архива дигиталних докумената. Наглашена је велика улога библиотекара научних библиотека у изградњи дигиталних репозиторија.

Кључне речи: Дигитални репозиторији, Покрет за отворене архиве, Dublin Core, OAI

OPEN ARCHIVE INITIATIVE (OAI)

Bogoljub Mazić
University Library "Svetozar Marković"
Belgrade

Abstract

The principles of the activities and main concepts for implementation of the open archives for digital documents are presented and defined. The importance of the role of the librarians from the scientific libraries in the digital repository development is stressed.

Key words: Digital repositories, Dublin Core, OAI, Open Archive Initiative

УВОД

Појам отворених архива све чешће се среће у литератури која се бави научним информацијама. У свету се одржавају и редовни скупови и конференције присталица и промотера овог покрета, као што су на пример годишње конференције посвећене електронским тезама и дисертацијама и скупови о научним комуникацијама и отвореним архивима у CERN-у у Женеви, на универзитету у Лунду итд. Корени покрета за отворене архиве - OAI (Open Archives Initiative) налазе се у групама које су се бавиле проблемима Е-издаваштва и које су прве промовисале и одржавале архиве научних радова на мрежи ради ширег приступа резултатима научног рада.

Почетна намера у покрету OAI била је мотивисана жељом да се развије оквир за међусобну разменљивост докумената ради формирања архива електронских издања. Убрзо је постало јасно да су циљеви промовисани у покрету OAI, као што су откривање метаподатака у различитим структурама помоћу јединственог протокола за његово прикупљање, превазишли почетни интерес Е-издаваштва. Услед тога је покрет проширио своју мисију на приступ свим типовима дигиталних докумената и за свој циљ одредио развој и промовисање стандарда за међусобну разменљивост електронских докумената у сврху ефикасне дисеминације њиховог садржаја.

¹ Рад је саопштен на Стручном скупу "Образовање кадрова и корисника у библиотечко-информационој делатности" у оквиру 10. скупштине Заједнице библиотека универзитета у Србији, који су одржани 25. октобра 2004. г. у Народној библиотеци Србије, Београд.

¹ Paper presented on the Professional Meeting "Education of library and information professionals and users" organized by the Serbian Academic Library Association and held 25 October 2004 in the National Library of Serbia, Belgrade.

Термин “архив” у називу OAI произилази из његовог порекла у Е-издаваштву где је израз архив генерално прихваћен као синоним за депозитну збирку тј. репозиториј научних радова. Архивисти међутим са правом примећују да се овај приступ разликује од њиховог изворног приступа који подразумева дуготрајно чување, заштиту, обраду и институционализацију архивске градје. OAI користи термин архив у ширем смислу : као назив за депозитну збирку похрањених информација. Како језик и термини нису никад једнозначни и недвосмислени, покрет OAI очекује да ће архивисти имати разумевања за коришћење израза архив и у овом ширем смислу.

Друга реч из назива – “отворен”, користи се из перспективе архитектуре система и дефинише и промовише рачунарске интерфејсе који чине садржаје различитих провајдера приступачним. При томе отвореност не значи слободан и неограничен приступ депозитним збиркама.

Учешће у OAI могуће је на два различита начина: у својству провајдера података и у својству провајдера сервиса. Некада обе ове улоге може имати један исти провајдер.

Провајдер података одржава једну или више депозитних збирки односно веб сервера које су подржане OAI протоколом да би постале доступне ширем корисничком кругу.

Провајдер сервиса шаље захтеве серверима провајдера података преко OAI-PMH протокола ради прикупљање метаподатака и помоћу њих гради основу са којом својим сервисима додаје нову вредност.

Обе врсте провајдера могу да се региструју на интернет страницама OAI и тиме огласе да су прихватили OAI протокол. Мотивација провајдера која их наводи да прихвате OAI протокол зависи од врсте учешћа у покрету OAI. Код провајдера података прихватање тог протокола је мотивисано учешћем у заједници која је колективно одлучила да користи OAI протокол као начин боље промоције садржаја које је креирала. Они такође могу бити заинтересовани и за функционалност који пружају поједини провајдери сервиса, као што њихов интерес може да произилази и из жеље за заједничким коришћењем знања. Провајдери сервиса прихватају овај протокол јер им он пружа могућност да своје сервисе обогате и метаподацима добијеним од различитих провајдера података.

Иако мисија OAI није ограничена само на метаподатке, њен превасходни циљ је да дефинише најшири оквир разменљивости међу депозитним збиркама уз најнижи степен ограничења, што је најлакше управо са метаподацима. Протокол за прикупљање метаподатака (OAI-PMH) дефинише механизам за прикупљање XML форматираних метаподатака из депозитних збирки али не и механизме за прикупљање података у другим форматима. Протокол такође не прописује начин повезивања метаподатака са његовим садржајем. Како многи корисници желе да сазнају садржај прикупљених података провајдери могу у самом метаподатку да дефинишу линк на опис садржаја податка. За опис садржаја се користи Dublin Core формат који обезбеђује довољан број елемената за препознавање садржаја.

Технички оквир покрета OAI је намерно једноставан како би учесницима обезбедио најнижи степен ограничења за пренос података. Сродни протоколи као што је нпр. 339.50 су потпуније функционалности и баве се управљањем упитима и обрађивањем добијених резултата уз помоћ предиката којима се враћени записи филтрирају. Међутим, оваква функционалност повећава трошкове имплементације и коришћења. Технички оквир покрета OAI не намерава да замени друге протоколе већ да понуди алтернативу за лакшу имплементацију и коришћење ради другачијих циљева од оних које постојећа решења за међусобну разменљивост података нуде. Даље коришћење OAI протокола ће показати да ли је оваква интероперабилност са ниским степеном ограничења реална и функционална, али судећи по томе колико овај покрет придобија нових присталица сваке године може се закључити да је покрет успео у својој мисији.

Опис OAI-PMH протокола - протокола за прикупљање података

- OAI-PMH протокол је базиран на HTTP протоколу за пренос података
- захтева аргументе као што су GET или POST.
- подржава шест врста захтева
- одговори су кодирани у XML синтакси
- може да подржи било који формат за метаподатке о садржају, а свакако Dublin Core

- депозитне збирке се логично деле у хијерахијски структуриране делове
- у току преноса података обликују се поруке о грешкама и
- контролише пренос података.

Протокол се заснива на следећим кључним појмовима:

Прикупљач (harvester) је апликација на рачунару клијента који шаље ОАИ-РМН захтев.

Резоизаториј (депозитна збирка) је сервер на мрежи способан да правилно обради ОАИ-РМН захтеве.

Ресурс (извор података) је објект који метаподатак описује, при чему природа извора податка није дефинисана ОАИ-РМН протоколом.

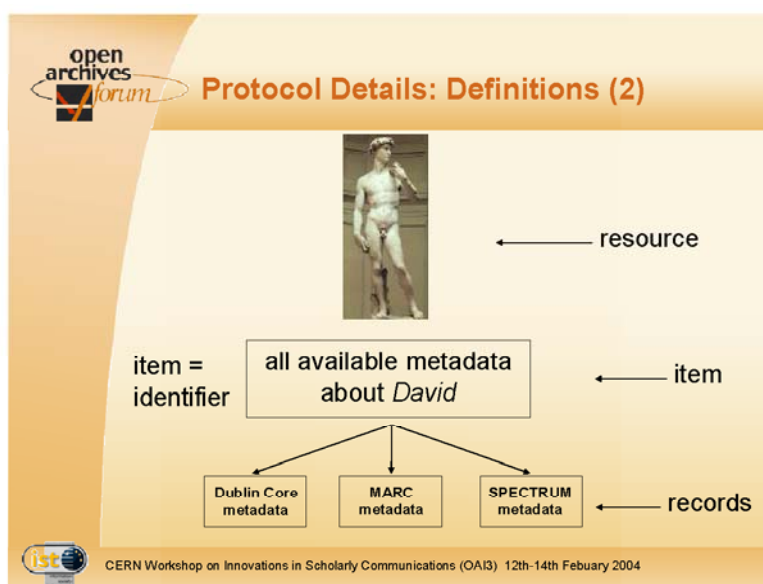
Јединица (item) је саставни део збирке из које се метаподатак преноси (дисеминира).

Запис (record) је метаподатак у одређеном формату.

Идентификатор (identifier) је јединствени кључ јединице у збирци.

Скуп (set) је опциони облик груписања јединица у збирци.

Пример за објашњење ових појмова у случају Микеланђеловог Давида:



Запис метаподатака се састоји из три дела :

заглавља (header) који је обавезан и садржи идентификатор и ознаку датума, самог метаподатка који је такође обавезан и кога чине XML кодирани подаци и пратеће информације (about) који је необавезан део и садржи податке о правима коришћења и пореклу.


Протокол подржава пренос и метаподатка у којем је опис садржаја метаподатка. Збирке података морају бити оспособљене за дисеминацију записа у Dublin Core формату. Тај формат се састоји од 15 елемената који су сви опциони и могуће их је понављати. Ти елементи садрже следеће податке : наслов, доприносилац (аутор), извор података, креатор, датум, језик, предмет, тип документа, сродне јединице, опис јединице, формат, покривеност, подаци о издавачу, идентификатор и подаци о ауторским правима.

Током неколико последњих година развоја, Dublin Core је постао стандард за метаподатке готово свих стручних дисциплина и стога се сматра одговарајућим избором у сваком скупу метаподатака.

Захтеве које шаље прикупљач (harvester) могу бити формулисани на 6 различитих начина при чему он не мора све користити али сервер депозитне збирке мора имплементирати све видове захтева.

Example: http://edoc.hu-berlin.de/OAI-2.0?verb=ListIdentifiers&from=2002-01-06&until=2002-01-08&metadataPrefix=oai_dc&set=doctypes:dissertations

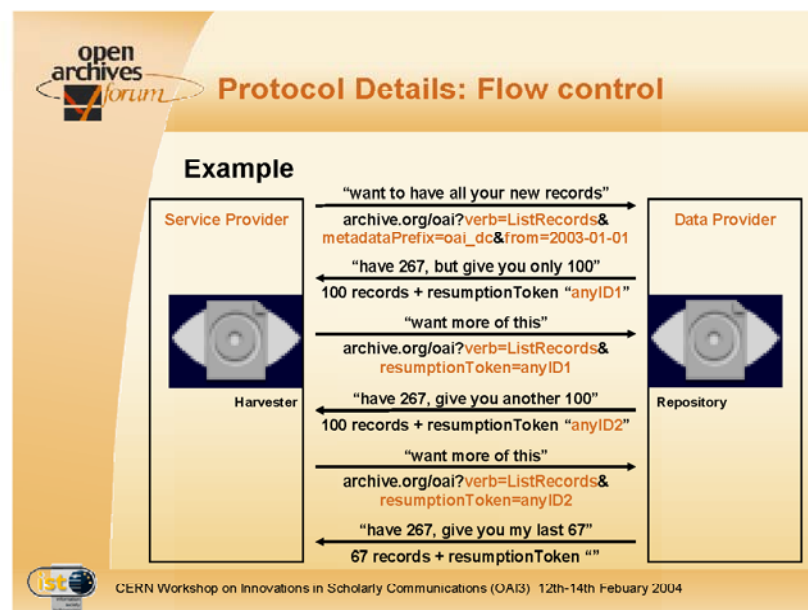
```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-10-22T17:49:49+01:00</responseDate>
  <request verb="ListIdentifiers" from="2002-01-03" until="2002-01-08" metadataPrefix="oai_dc"
    set="doctypes:dissertations">http://edoc.hu-berlin.de/OAI-2.0</request>
  <ListIdentifiers>
    <header>
      <identifier>oai:HU.Berlin.de:3000819</identifier>
      <datestamp>2002-01-08</datestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb33</setSpec>
    </header>
    <header>
      <identifier>oai:HU.Berlin.de:3000831</identifier>
      <datestamp>2002-01-07</datestamp>
      <setSpec>doctypes</setSpec>
      <setSpec>doctypes:dissertations</setSpec>
      <setSpec>dnb</setSpec>
      <setSpec>dnb:dnb27</setSpec>
    </header>
  </ListIdentifiers>
</OAI-PMH>
```

 CERN Workshop on Innovations in Scholarly Communications (OAI3) 12th-14th February 2004

Подела депозитних збирки на логичке делове је опциона, архиви нису обавезни да то чине и за њих не постоје предвиђене препоруке. Међутим, подела на логичне целине је веома корисна те је стога најчешће одређују заједнице којима су те збирке намењене. Најчешће подела је према типовима публикација на тезе, чланке и слично, према врсти грађе на текстуалне записе, аудио записе, слике и итд. и према садржају на научне дисциплине на нпр. биологију, медицину, историју, итд.

Ознака датума садржи датум задње измене у скупу метаподатака и обавезан је за сваку јединицу.

Наредни дијаграм приказује проток података које подржава OAI-протокол.



Полазишта за успешну имплементацију OAI

Пре успостављања OAI система провајдер података и провајдер сервиса морају да се договоре око кључних питања за функционисање овог система. Провајдер података треба да зна које податке жели да пружи и које провајдере сервиса жели да снабдева овим подацима. Провајдер сервиса мора да зна које сервисе жели да нуди, од којих провајдера података жели да преузима метаподатке и на који начин је потребно да су ти метаподаци обрађени. Успешан договор провајдера око ових питања је први корак ка успостављању OAI система.

За рад са метаподацима потребно је следеће :

Dublin Core формат,
дефиниција XML схеме,
опис извора података и
дефинисана хијерархијска структура збирке која ће омогућити селективно прикупљање података.

Провајдер података мора да задовољи следеће претходне услове :


- Метаподатке чува у CQL базама података а по потреби и у фајл систем датотекама; свака јединица метаподатка треба да садржи јединствену идентификациону ознаку,
- Метаподаци се налазе на веб серверима (апацхе или ИИС) којима се приступа преко Интернета,
- Сервер провајдера података мора да има развијен Веб АПИ сервис који омогућава да провајдер сервиса лако нађе и обради информације,
- Идентификатор архива базира се на УРЛ,
- Опис садржаја метаподатка потребан је у Dublin Core формату и
- Већи архиви имплементирају и контролу протока података.

Провајдер сервиса мора такође да задовољи један низ претходних услова :

- Обезбеди сервере који су на Интернету,
- Обезбеди програме за управљање базама података који су релациони или XML,
- Обезбеди програмско окружење које ће омогућити да се коришћењем HTTP протокола шаљу захтеви различитим веб серверима,
- Обезбеди подршку за слање захтева базама података и
- XML парсер.


Запис метаподатка треба да задовољи следеће основне критеријуме :

- протокол OAI-PMH прописује коришћење XML схеме која нам указује шта је дозвољено у XML документу и састоји се од ознаке имена и ознаке физичке локације. Било који XML који има XML схему добар је за OAI.
- протокол OAI-PMH прописује коришћење "OAI_DC" схеме
- OAI-PMH документација садржи и схеме следећих метаподатака: RFC 1807 метадата, MARC21 метадата и OAI_DC метадата



Example: http://edoc.hu-berlin.de/OAI-2.0?verb=GetRecord&identifier=oai:HUBerlin.de:3000819&metadataPrefix=oai_dc

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-11-27T14:37:01+01:00</responseDate>
  <request verb="GetRecord" metadataPrefix="oai_dc"
    identifier="oai:HUBerlin.de:3000819">http://edoc.hu-berlin.de/OAI-2.0</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:HUBerlin.de:3000819</identifier>
        [...]
      </header>
      <metadata>
        <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
          xmlns:dc="http://purl.org/dc/elements/1.1/"
          xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
            http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
          <dc:title>Einfluss genetischer Variationen im Tumor Nekrose [...]</dc:title>
          <dc:creator>Schäffler, Antje</dc:creator>
          [...]
        </oai_dc:dc>
      </record>
    </GetRecord>
  </OAI-PMH>
```



CERN Workshop on Innovations in Scholarly Communications (OAI3) 12th-14th February 2004

“ОАИ_ДЦ” схема је намењена метаподацима који описују садржај метаподатка и за њих се користи неадресована Dublin Coreсхема. Уз Dublin Coreсхему обавезно је користити и “контејнер” схеме и оне се могу добити на ОАИ сајту.

Поред ОАИ_ДЦ формата за метаподатке могуће је имплементирати и друге податке као што је ИМС/ИЕЕЕ ЛОМ или било који други МЕС уколико је XML кодиран и има XML схему.

Избор софтверских алата зависи од депозитне збирке и од сервиса. Већина ових програма је бесплатна и могуће их је дограђивати. Такође су доступни и за различите оперативне системе а разликују се према нагласку који формат метаподатака подржавају, каква им је могућност конфигурисања и да ли имају програмске библиотеке. До исцрпне листе ових програма долази се преко адресе : [HTTP://www.soros.org/openaccess/software/](http://www.soros.org/openaccess/software/)

За проналажење провајдера података и рекламирање свог сервиса препоручују се адресе на следећој слици:



How to advertise your service and find data providers

- Repository Explorer
<http://oai.dlib.vt.edu/cgi-bin/Explorer/oai2.0/testoai>
- OAISTER
<http://www.oaister.org/o/oaister/>
- Southampton
<http://archives.eprints.org/eprints.php>



CERN Workshop on Innovations in Scholarly Communications (OAI3) 12th-14th February 2004

Закључак

У последњих десет година се припрема за штампу обавља углавном на рачунарима али после тога се ти електронски документи не похрањују систематски, већ се губе без трага, односно остају само евентуално на приватним рачунарима својих аутора. С обзиром на то да се све већи део научне продукције и не штампа него остаје само у дигиталном облику, веома је важно да се та чињеница има у виду при планирању класичне библиотечке делатности – прикупљања, чувања и давања на коришћење библиотечког материјала.

Тренд електронских архива је већ више година јако присутан у свету, те и у Србији морамо озбиљно прићи решавању проблема архивирања електронских докумената на систематски начин. У томе нам значајно може помоћи и слободно доступан водич [Guide to Institutional Repository Software](#). Водич описује пет тренутно доступних, ОАИ прилагођених система који су засновани на Open Source принципу.

Оно што је веома важно имати у виду је да су у успостављању мреже отворених архива технички проблеми много лакше решиви од организационих. Библиотечка заједница треба да се припреми да заузме своје место у области електронских научних комуникација, преузимањем на себе задатка обраде, чувања и обезбеђивања коришћења електронских докумената. Иако су стандарди и начин руковања документима промењени, то је и даље посао библиотекара.

Оно што је веома важно имати у виду је да су у успостављању мреже отворених архива технички проблеми много лакше решиви од организационих. Библиотечка заједница треба да се припреми да заузме своје место у области електронских научних комуникација, преузимањем на себе задатка обраде, чувања и обезбеђивања коришћења електронских докумената. Иако су стандарди и начин руковања документима промењени, то је и даље посао библиотекара.

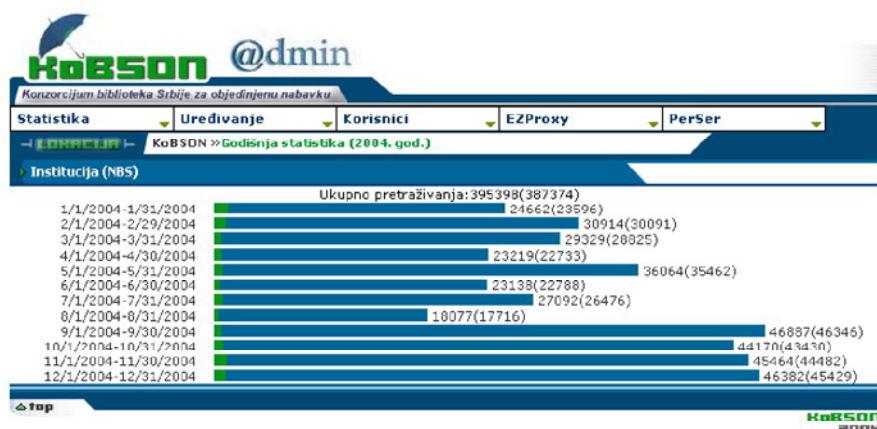
KoBSON

I ove godine Ministarstvo nauke i zaštite životne sredine nastavlja sa finansiranjem objedinjene pretplate stranih baza podataka i časopisa u papirnoj formi. Sve informacije o dostupnim bazama, časopisima i uslugama možete naći na:

<http://nainfo.nbs.bg.ac/KoBSON>

Novi servisi:

- kako i gde dobiti podatke o citiranosti istraživača
- kako i gde obaviti međubibliotečku pozajmicu
- detaljne statistike o broju preuzetih članaka za 2004. godinu po servisima
- detaljnu informaciju o časopisima u papirnoj formi
- radovi "naših" istraživača u Web of Science
- pristup do najznačajnijih svetskih otvorenih arhiva
- sve informacije o licenci za pristup proxy serveru
- desetominutno uputstvo o pretraživanju kataloga elektronskih časopisa



Uspešno pretraživanje i puno pogodaka,
Vaš

